

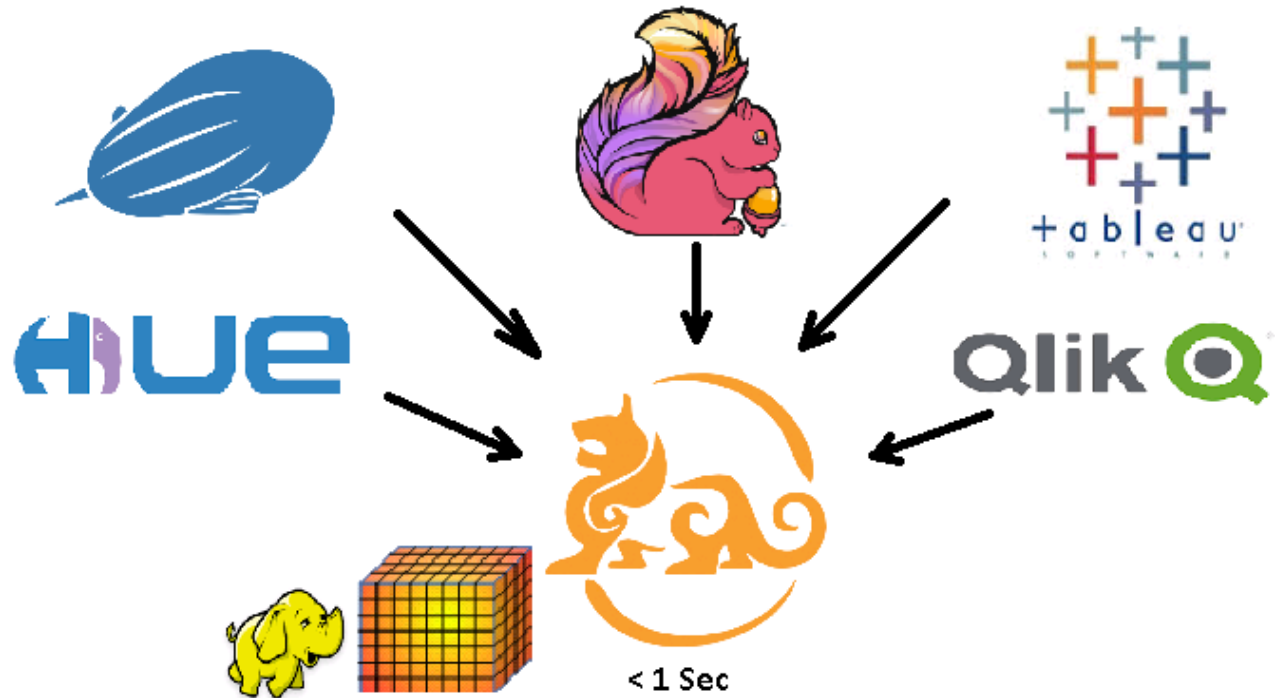
BI Tools and Apache Kylin

Date: Sep 2016

Updated: Dec 2016

Author: Ramón Portolés, Alberto a.ramonportoles@gmail.com

[Linkedin](#)



Intro: Kylin Project

[Apache Kylin](#) is a powerful software which enables Extrem OLAP cubes on Hadoop system, with lineal scalability and sub-second response time over TB of data

Mainly developed by [eBay](#) and [Kyligence.io](#) teams

Unfortunately is little known, perhaps some causes can be:

- Few documentation: some nice features are undocumented, only developers know how to use them
- It is developed / used by big companies with many technical and human resources

- The access to data, is using a integrate UI Kylin or ODBC / API Rest, too complex for most of users or PoCs

Intro: Integration Study with BI

During two months (July & August 2016) I tried to connect kylin with most popular BI Tools, sometimes has been successful others it hasn't















Officially the [Kylin ODBC Driver](#) only has been tested with [Tableau Software](#)

We evaluate two aspect:

- Connectivity to import data
- Integration of this data with different features of this BI Tool

For our test use : Use [Kylin v1.5.2](#) (on docker) [Kylin ODBC v1.5](#)

Resume Table

Soft	Data Import	Data Integration	Detailed Manual	Version	License
Tableau Desktop			GitHub	v9.3.3	Free
Apache HUE			GitHub	v3.10	GPL
Qlik Desktop			GitHub	v3.0.2.0	Free
Apache Zeppelin			External	V0.7	GPL
Apache Flink			GitHub	V1.2	GPL
SQuirreL			GitHub	3.7.1	GPL
Kylin Caravel			Internal*	v0.91	GPL
Sisense			Internal*	v6.4.1	Demo

Microsoft PowerBI	☆☆☆		N.A	v2.38.4491	Free
Dundas BI	☆☆☆		Intenal*	V3.0.2.100	Demo
Microsoft SSRS: Report Builder	☆☆☆		Internal*	V2012 SP3	Demo
Microsoft SSRS: SSDT	☆☆☆		Internal*	V2012 SP3	Demo

Internal*: Despite efforts, any remarkable result ☹️, But if you are interested I can share them with you

License:

- GPL: Full features are free , code free
- Free: With no code, but for free.
- Demo: Limited usage to certain period of time.

Tableau

[View full Report](#)

Resume:

- Some warning importing metadata but no issues when it is done.
- The data integration is complete

Apache Hue

[View full Report](#)

Resume:

- Some issues and code changes must be made before compiling (see steps in manual)
- View / Import meta-data is not available, more changes in code are needed(exits GIRA)
- Only 1000 rows can be Imported (exits GIRA)
- Isn't compatible with Dashboards (In Road-map for v4)
- It's a dynamic and important open source project, in next versions > 3.11 some bugs will be fixed

Qlik Desktop

[View full Report](#)

Resume:

- Because there isn't an official connector, you must write "load script", it's easy and works fine (see instructions)
- This type of data have some limits:
 - You can't see the data preview
 - The associations (between tables are automatically made), can't be change. This means if they aren't correct, your results will be wrong
- [Technical Support](#)

SquirrelL (Unversal SQL Client)

[View full Report](#)

The configuration is not very intuitive, but works very well and get us a lot of information about tables metadata

Indicated for PoC or Kylin UI alternative

Not use ODBC driver uses JDBC directly then:

- Good for performance and avoid ODBC problems
- The version of Kylin and JDBC must be the same (upgrade Kylin implies upgrade your Java lib)

Apache Zeppelin

[Apache Kylin Interpreter for Apache Zeppelin](#)

[Apache Zeppelin](#) is Web notebook very useful for generate dynamic reports which integrate data from differents datasources (Hive, Spark, Kylin, Flink, PostgreSQL, HBase, ..)

You can found exaples in [ZeppelinHub](#) and [Hortonworks Zeppelin Gallery](#)

Apache Flink

[ReadKylinFromFlink](#)

[Apache Flink](#) is a Framework to process data in batch or streaming mode

Now we can read Kylin's data from Apache Flink, Flink 1.2 or higher is recommended

Kylin Caravel

[Caravel](#) is a UI for data exploration, developed by [Airbnb](#), that uses [Druid](#), as database
[Rocky Qi](#) (thanks, for the suggestions) creates a [fork Kylin – Caravel](#), to support Kylin as database

After 2 weeks of running tests, (there is a bug in the internal metadata database) and previous branch didn't work on Ubuntu (there is some problems with this SO) and Centos... I desisted (sorry)

I will retry in the future / next version. Its a good candidate for docker (and makes life easier for others) (I like docker !!!)

Sisense

The answer of technical support of Sisense was that is compatible with Kylin

But, after contact with [technical support](#), and some emails: it didn't work and there was no manual that could explain how to configure it to make it work.

You can see [this video](#)

Microsoft PowerBI Desktop

I tried to import data using the automatic mode and writing select queries. In the best case, I could see the preview of data before an import. But the import never worked.

In logs, I saw some data conversion errors / overflows. You can see [this video](#)

The problem are some issues of ODBC connector :

- [KYLIN-2121](#) v1.6.0 Bug "Value was either too large or too small for an Int32"
- [Kylin 2235](#) vx.x Bug "array dimensions exceed supported ranges"

Dundas BI

(Thanks to [Ariel Pohoryles](#), of Dundas team)

Dundas BI, Was one of the first to do graphs with high visual quality

Microsoft purchased a license to use Dundas Chart on SQL Reporting services 2008 and above

There is a demo for 45 days

Nowadays, there is a bug reading metadata ([Kylin 2274](#)), After review problem with Dudas Team sound like a problem with ODBC ([explanation](#))

Microsoft SSRS: Report Builder

Microsoft Report builder , is a end user tool for create your personal Reports (from Web UI) on SSRS server or SharePoint

Its free with Microsoft SQL Server license

You can define Kylin as DataSource, but when use it to create DataSet fail: “Data Source name not found and no default driver speficied”

Microsoft SSRS: Visual Studio Data Tools (SSDT)

Is an extension to generate professional reports of SSRS using Visual Studio

Its free with Microsoft SQL Server license

The error is similar to that Report Builder “Data Source name not found and no default driver speficied”

It seems that only work with certains providers proposed by default by Microsoft

Others

Other tools that nowadays, can't work with Apache Kylin (need changes in code, doesn't support SQL, doesn't support ODBC sources, ...)

[Kibi](#) /[Kibana](#)

From: [Giovanni Tummarello](#) (thanks, for the info)

“Kibi dashboard DO NOT support SQL or JDBC a Kibi datasource is only used for special filters or certain other operations, check the documentation. We will likely support it in the future though. saying this so that you dont waste your time.. you will not be able to see anything now with Kibi if you dont load things in ES also”

Ploty

From: Chris Parmer (thanks, for the info)

“Thanks for reaching out! We haven't added this yet. Do you use Apache Kylin at your work? We are going to be adding spark and redshift next before we get to something like Kylin.”

Dataiku

(Thanks to [Hugo Schmitt](#))

It's a Visualization and Machine Learning tool. Imagine, if you have sales cube by dates, predict the sales using data of previous years ;)

Extensible using [Plugins](#)

Dataiku, don't support standard ODBC connector, the workaround is test with unofficial JDBC support

ToDo:

Test and doc the integration of Kylin with [Saiku](#)

For any suggestions, feel free to contact me

Thanks, Alberto